Asymptotics of set-indexed conditional empirical processes based on dependent data *

Wolfgang Polonik
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg, Germany

Qiwei Yao
Institute of Mathematics and Statistics
University of Kent at Canterbury
Canterbury, Kent CT2 7NF, UK

Abstract

We study the asymptotic properties of the conditional empirical process based on

$$\widehat{F}_n(C|x) = \sum_{t=1}^{n} \omega_n(X_t - x) I_{\{Y_t \in C\}}$$

indexed by $C \in \mathcal{C}$, where $\{(X_t, Y_t), t = 1, \ldots, n\}$ are observations from a strong mixing stochastic process, $\{\omega_n(X_t - x)\}$ denote some kernel weights, and $\mathcal{C}$ is a class of sets. Under the assumption on the richness of the index class $\mathcal{C}$ in terms of metric entropy with bracketing, we have established uniform convergence, and asymptotic normality for $\widehat{F}_n(\cdot|x)$. The key technical result gives rates of convergences for the sup-norm of the conditional empirical process over a sequence of classes $\mathcal{C}$ with decreasing maximum L_1-norm. The results are then applied to derive Bahadur-Kiefer type approximations for a generalized conditional quantile process which is closely related to the minimum volume sets. The potential applications in the areas of estimation of level sets and testing for unimodality of conditional distributions are discussed.

Keywords: Bahadur-Kiefer approximation, conditional distribution, covering number, empirical process theory, generalized conditional quantile, level set, minimum volume predictor, Nadaraya-Watson regression estimator, nonlinear time series, strong mixing.

*Supported partially by the EPSRC Grants L67561 and L16385, and the EU Human Capital and Mobility Program ERB CHRX-CT 940693.

1 Introduction

An empirical process indexed by a class of sets or functions is an interesting mathematical model with various statistical applications (see, for example, Shorack and Wellner 1986). Such a process defined from independent and identically distributed (*i.i.d.*) observations has been extensively studied in literature in past two decades. More recently, empirical processes based on dependent data have been studied under various mixing conditions (*e.g.* Massart 1987, Andrews and Pollard 1994, Doukhan, Massart, and Rio 1995). The extension of the above exploration to *conditional* empirical processes is practically useful and certainly more technically challenging. To our knowledge, the study so far has been confined to the cases with *i.i.d.* observations, which includes, among others, Stute (1986a,b), Horvath (1988), and Bhattacharya and Gangopadhyay (1990). Conditional quantile processes, which will also play a role in the present paper, are closely related to conditional empirical processes; for the *i.i.d.* case see, for example, Bhattacharya and Gangopadhyay (1990), Mehra et al. (1991), Gangopadhyay and Sen (1993), and Xiang (1995, 1996).

In this paper we study asymptotic properties of set-indexed conditional empirical processes based on observations from stochastic processes which are strong mixing. The Bahadur-Kiefer type approximations for a generalized conditional quantile process are derived, which has a direct bearing on asymptotic properties of minimum volume sets. Although our study is directly motivated by prediction of nonlinear and non-Gaussian time series (Polonik and Yao 1998), we also briefly discuss potential applications of these results in various other statistical practices such as level set estimation and testing for unimodality.

Let $\{(X_t, Y_t)\}$ be a strictly stationary process, with $X_t \in \mathbf{R}^d$ and $Y_t \in \mathbf{R}^{d'}$. Let $F(\cdot|x)$ be the conditional distribution of Y_t given $X_t = x$. Note that for any given measurable set $C \subset \mathbf{R}$, $E\{I_{(Y \in C)}|X_t = x\} = F(C|x)$. This regression relationship suggests the following Nadaraya-Watson estimator for $F(\cdot|x)$ from the observations $\{(X_t, Y_t),\, t = 1, \ldots, n\}$:

$$\widehat{F}_n(C|x) = \sum_{t=1}^{n} I_{\{Y_t \in C\}} K\left(\frac{X_t - x}{h}\right) \Big/ \sum_{t=1}^{n} K\left(\frac{X_t - x}{h}\right), \tag{1.1}$$

where $K(\cdot) \geq 0$ is a kernel function on $\mathbf{R}^d$, and $h > 0$ is a bandwidth. $\widehat{F}_n(\cdot|x)$ is called *empirical conditional distribution*. In the usual time series context, Y_t is a scalar and X_t consists of its lagged values. To predict Y_t from X_t for non-Gaussian time series data, conventional interval predictors such as the mean plus and minus a multiple of standard deviation are no longer pertinent. The conditional minimum volume (*i.e.* Lebesgue measure) predictor could perform substantially better than the conditional quantile interval. The estimation of the conditional distribution $F(C|X_t = x)$

for $C \in \mathcal{C}$, where $\mathcal{C}$ is an appropriate class of measurable sets, plays a key role in deriving minimum volume predictors (see Polonik and Yao 1998).

The goal of this paper is to study the asymptotic behaviour of the conditional empirical process

$$\mathcal{C} \ni C \to \nu_n(C|x) = \sqrt{nh^d}\{\widehat{F}_n(C|x) - F(C|x)\}. \tag{1.2}$$

The main result Theorem 2.3 essentially deals with the asymptotic behaviour of the modulus of continuity of $\nu_n(\cdot|x)$. It turns out that this asymptotic behaviour depends on the richness (or complexity) of the index class $\mathcal{C}$, which is measured in terms of metric entropy with bracketing. In fact if $\mathcal{C}$ is not too rich (see Theorem 2.3 below), the conditional $\mathcal{C}$-indexed empirical process converges weakly, in the sense of Hoffman-Jørgensen (*cf.* van der Vaart and Wellner 1996), to a so-called $F(\cdot|x)$-bridge. This means that the empirical process behaves like the one based on *i.i.d.* observations in term of first order asymptotics, as long as the class $\mathcal{C}$ is not too rich. This phenomenon is not unexpected, since only the observations with X_t in a small neighbourhood of x are effectively used in the estimation (1.1). Those observations are not necessarily close with each other in the time space. Indeed, they could be regarded as asymptotically independent under appropriate conditions such as strong mixing.

On the other hand, it remains at least to us as an open problem to identify the maximum richness of $\mathcal{C}$ (under the strong mixing condition) to retain the above *i.i.d.*-like asymptotic behaviour. The condition specified in this paper restrains $\mathcal{C}$ far from being as rich as in the case of *i.i.d.* observations in order to retain the same asymptotic results. Note that the standard conditional empirical processes indexed by $x \in \mathbf{R}^d$ usually behave asymptotically like those based on *i.i.d.* observations. However, the corresponding class $\mathcal{C} = \{(-\infty, x], x \in \mathbf{R}^d\}$ is very "thin".

As mentioned above, a (generalized) conditional quantile process also plays a crucial role in the present paper. To this end, we introduce the notion of minimum volume sets (MV-sets) first. For $\alpha \in [0,1]$, the set $\mathrm{M}_{\mathcal{C}}(\alpha|x) \in \mathcal{C}$ satisfying the condition that

$$\mathrm{M}_{\mathcal{C}}(\alpha|x) \in \mathrm{argmax}\{\mathrm{Leb}(C) : C \in \mathcal{C}, F(C|x) \geq \alpha\} \tag{1.3}$$

is called a *conditional MV-set* in $\mathcal{C}$ at level α, where $\mathrm{Leb}(\cdot)$ denotes Lebesgue measure. Analogously, $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ denotes an *empirical* conditional MV-set if $F(\cdot|x)$ in (1.3) is replaced by the empirical distribution $\widehat{F}_n(\cdot|x)$. We denote their volumes as

$$\mu_{\mathcal{C}}(\alpha|x) = \mathrm{Leb}(\mathrm{M}_{\mathcal{C}}(\alpha|x)) \quad \text{and} \quad \widehat{\mu}_{\mathcal{C}}(\alpha|x) = \mathrm{Leb}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)), \tag{1.4}$$

respectively. The volume process

$$\alpha \to \sqrt{nh^d}\ (\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x)) \tag{1.5}$$

can be considered as a conditional version of a generalized quantile process as defined in Einmahl and Mason (1992). For this process a Bahadur-Kiefer-type approximation is given in §3, which, in the special case of $d = 1$ and the observations being independent, improves the result of Bhattacharya and Gangopadhyay (1990). Polonik (1997) established similar results for an unconditional volume process based on *i.i.d.* observations.

The rest of the paper is organized as follows. We present the asymptotic results of the process $\{\nu_n(\cdot|x)\}$ in §2. §3 contains the Bahadur-Kiefer approximations for the volume process (1.5). §4 provides a brief discussion on how the results in this paper can be applied to various statistical applications. §5 contains all the technical proofs.

2 The set-indexed conditional empirical process

In this section, we establish asymptotic properties of the process $\nu_n(\cdot|x)$ defined in (1.2), which include a Glivenko-Cantelli type result, the asymptotic normality of finite dimensional distributions, and the asymptotic behaviour of the modulus of continuity. The two latter imply that $\nu_n(\cdot|x)$ converges to a Gaussian process.

Let $f(\cdot)$ be the density function of X_t. We always assume that $x \in \mathbf{R}^d$ is fixed and $f(x) > 0$. Further, all the non-deterministic quantities are assumed to be measurable, and we write $d_{F(\cdot|x)}(A, B) = F(A \Delta B|x)$. We use c to denote some generic constant, which may be different at different places. We introduce some regularity conditions first.

(A1) The marginal density f is bounded and continuous in a neighbourhood of x.

(A2) The kernel density function K is bounded and symmetric, and $\lim_{u\to\infty} ||u||^d K(u) = 0$.

(A3) $f \in C_{2,d}(b)$, where $C_{2,d}(b)$ denotes the class of bounded real-valued functions with bounded second order partial derivatives.

(A4) $F(\cdot|x)$ has a Lebesgue-density $g(\cdot|x) \in C_{2,d'}(b)$. Moreover, for each $C \in \mathcal{C}$ the function $F(C|\cdot) \in C_{2,d}(b)$ such that $\sup_{C\in\mathcal{C}} \left|\frac{\partial^2}{\partial x_i \partial x_j} F(C|x)\right| < \infty, \ \forall\, 1 \le i, j \le d$.

(A5) $|| \int vv^T K(v)\, dv|| < \infty$.

(B1) The joint distribution of (X_t, X_{t+q}) has the density function f_q, and $\sup_{q\ge 1} ||f_q||_p < \infty$ for some $p \in (2, \infty]$.

(B2) The joint density function of (X_s, X_t, X_q, X_r) exists and is bounded from the above by a constant independent of (s, t, q, r).

We call the stationary process $\{(X_t, Y_t)\}$ *strong mixing* if

$$\alpha(j) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_j^\infty} |P(AB) - P(A)P(B)| \to 0, \quad \text{as } j \to \infty, \tag{2.1}$$

where $\mathcal{F}_s^t$ denotes the σ-algebra generated by $\{(X_i, Y_i), s \le i \le t\}$. We use the term *geometrically* strong mixing if $\alpha(j) \le a\, j^{-\beta}$ for some $a > 0$ and $\beta > 1$, and *exponentially* strong mixing if $\alpha(k) \le b\, \gamma^k$ for some $b > 0$ and $0 < \gamma < 1$. Sometimes the condition of strong mixing can be reduced to so-called 2-*strong mixing*, which is defined as in (2.1) with $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_j^\infty$ replaced by $\sigma(X_0, Y_0)$ and $\sigma(X_j, Y_j)$ respectively. We use the terms *geometrically* or *exponentially* 2-strong mixing in the similar manners.

Now we introduce the notion of *metric entropy with bracketing* which provides a measure of richness (or complexity) of a class of sets $\mathcal{C}$. This notion is closely related to *covering numbers*. We adopt L_1-type covering numbers using the bracketing idea. The bracketing reduces to *inclusion* when it is applied to classes of sets rather than classes of functions. For each $\epsilon > 0$, the covering number is defined as

$$\begin{aligned} N_I(\epsilon, \mathcal{C}, F(\cdot|x)) = \inf\{ & n \in \mathbf{N} : \exists\, C_1, \ldots, C_n \in \mathcal{C} \text{ such that} \\ & \forall\, C \in \mathcal{C} \;\; \exists\, 1 \le i, j \le n \text{ with } C_i \subset C \subset C_j \text{ and } F(C_j \setminus C_i | x) < \epsilon\}. \end{aligned} \tag{2.2}$$

The quantity $\log N_I(\epsilon, \mathcal{C}, F(\cdot|x))$ is called *metric entropy with inclusion* of $\mathcal{C}$ with respect to $F(\cdot|x)$. A pair of sets C_i, C_j is called a *bracket* for C. Estimates for such covering numbers are known for many classes. (See, *e.g.* Dudley 1984.) We will often assume below that either $\log N_I(\epsilon, \mathcal{C}, F(\cdot|x))$ or $N_I(\epsilon, \mathcal{C}, F(\cdot|x))$ behave like powers of ϵ^{-1} : We say that condition (R_γ) holds if

$$\log N_I(\epsilon, \mathcal{C}, F(\cdot|x)) < H_\gamma(\epsilon), \quad \text{for all } \epsilon > 0, \tag{R_γ}$$

where

$$H_\gamma(\epsilon) = \begin{cases} \log(A\, \epsilon^{-r}) & \text{if } \gamma = 0, \\ A\, \epsilon^{-\gamma} & \text{if } \gamma > 0, \end{cases} \tag{2.3}$$

for some constants $A, r > 0$. In fact condition (R_0) holds for intervals, rectangles, balls, ellipsoids, and for classes which are constructed from the above by performing set operations union, intersection and complement finitely many times. The classes of convex sets in $\mathbf{R}^d$ $(d \ge 2)$ fulfill condition (R_γ) with $\gamma = (d-1)/2$. This and other classes of sets satisfying (R_γ) with $\gamma \ge 0$ can be found in Dudley (1987).

Now we are ready to formulate the results on the uniform consistency and the (pointwise) asymptotic normality of $\nu_n(C|x)$.

Theorem 2.1 (Uniform consistency)

Suppose that conditions (A1), (A2) *and* (B1) *hold, and that* $\{(X_t, Y_t)\}$ *is geometrically* 2*-strong mixing with* $\beta > 2(p-1)/(p-2)$*. Let* $\mathcal{C}$ *be a class of measurable sets for which* $N_I(\epsilon, \mathcal{C}, F(\cdot|x)) < \infty$ *for any* $\epsilon > 0$*. Suppose further that* $\forall\, C \in \mathcal{C}$

$$|F(C|y)f(y) - F(C|x)f(x)| \to 0 \quad \text{as } y \to x. \tag{2.4}$$

If $nh^d \to \infty$ *and* $h \to 0$ *as* $n \to \infty$*, then*

$$\sup_{C\in\mathcal{C}} |\widehat{F}_n(C|x) - F(C|x)| \xrightarrow{P} 0.$$

Theorem 2.2 (Asymptotic normality)

Let (A2) – (A5) *and* (B2) *hold, and suppose that* (B1) *holds with* $p = \infty$*. Suppose further that the process* $\{(X_t, Y_t)\}$ *is geometrically strong mixing with* $\beta > 2$*. Let* $h = cn^{-\frac{1}{d+4}}(\log\log n)^{-1}$*. Then as* $n \to \infty$*, for* $m \geq 1$ *and* $C_1, \ldots, C_m \in \mathcal{C}$*,*

$$\{\nu_n(C_i|x);\ i = 1, \ldots, m\} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where $\Sigma = (\sigma_{i,j})_{i,j=1,\ldots,m}$*, and* $\sigma_{i,j} = \{F(C_i \cap C_j|x) - F(C_i|x)F(C_j|x)\} \int K^2/f(x)$*.*

In order to formulate the next theorem which provides the information on the asymptotic behaviour of the modulus of continuity (see remarks below), we need to introduce the following function

$$\Lambda_\gamma(\sigma^2, n) = \begin{cases} \sqrt{\sigma^2 \log \frac{1}{\sigma^2}} & \text{if } \gamma = 0, \\ \max\left((\sigma^2)^{\frac{1-\gamma}{2}}, \left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}} \right) & \text{if } \gamma > 0. \end{cases} \tag{2.5}$$

Theorem 2.3 *Suppose that* (A2) – (A5) *and* (B1) *hold, and the process* $\{(X_t, Y_t)\}$ *is exponentially strong mixing. For each* $\sigma^2 > 0$*, let* $\mathcal{C}_\sigma \subset \mathcal{C}$ *be a class of measurable sets with* $\sup_{C\in\mathcal{C}_\sigma} F(C|x) \leq \sigma^2 \leq 1$*, and suppose that* $\mathcal{C}$ *fulfills* (R_γ) *with some* $\gamma \geq 0$*. Further we assume that* $h^d \to 0$ *and* $nh^d \to \infty$ *as* $n \to \infty$ *such that*

$$nh^{d+4} \leq \left(\Lambda_\gamma(\sigma^2, n)\right)^2. \tag{2.6}$$

For $\gamma = 0$ *we assume in addition that* $\frac{nh^d\sigma^2 \log \frac{1}{\sigma^2}}{(\log n)^6} \to \infty$ *as* $n \to \infty$*. Then there exists a constant* $M > 0$ *such that* $\forall\, \epsilon > 0$*,*

$$P\left(\sup_{C\in\mathcal{C}_\sigma} |\nu_n(C|x)| \geq M\,\Lambda_\gamma(\sigma^2, n) \right) \leq \epsilon$$

for all sufficiently large n *and all* $\sigma^2 < \sigma_0^2$*, where* $\sigma_0^2 > 0$ *is a constant.*

Remark 2.4 (a) Note that Λ_n tends to zero as $n \to \infty$ provided $\gamma < 1/3$. In this case, Theorem 2.3 entails the tightness of the conditional set-indexed empirical process. To see this, note that trivially $\sup_{C,D \in \mathcal{C}} |\nu_n(C|x) - \nu_n(D|x)| \leq 2 \sup_{B \in \mathcal{C} \setminus \mathcal{C}} |\nu_n(B|x)|$ where $\mathcal{C} \setminus \mathcal{C} = \{C \setminus D, C, D \in \mathcal{C}\}$. Without the loss of generality, we may assume that $\emptyset \in \mathcal{C}$ such that $\mathcal{C} \subset \mathcal{C} \setminus \mathcal{C}$. Now, it is easy to see that $N_I(\epsilon, \mathcal{C}, F(\cdot|x)) \leq N_I(\epsilon, \mathcal{C} \setminus \mathcal{C}, F(\cdot|x)) \leq (N_I(\epsilon/2, \mathcal{C}, F(\cdot|x)))^2$. This implies that (R_γ) holds for $\mathcal{C}$ if and only if it holds for $\mathcal{C} \setminus \mathcal{C}$. Hence, an application of Theorem 2.3 to the class $\mathcal{C} \setminus \mathcal{C}$ together with Theorem 2.2 entails, by standard arguments, that the set-indexed process converges in distribution to a so-called $F(\cdot|x)$-bridge, provided $\gamma < 1/3$. An $F(\cdot|x)$-bridge is a Gaussian process with almost surely continuous sample paths and covariance structure as given in Theorem 2.2 (*e.g.* Pollard 1984). Taking into account possible non-measurability the convergence in distribution should be understood in the sense of Hoffman-Jørgensen (see van der Vaart and Wellner 1996).

(b) It is well-known in the empirical process theory that an unconditional empirical process based on *i.i.d.* observations is tight if (R_γ) holds with the sharp bound $\gamma < 1$ (see Alexander 1984). The same conclusion holds for a conditional empirical process as long as the process is formed from a set of *i.i.d.* observations. However, for the empirical processes based on dependent data under the strong mixing condition, we assume in this paper $\gamma < 1/3$ to achieve the tightness. It was indicated on page 128 of Andrews and Pollard (1994) that the tightness of an (unconditional) empirical process can be established by using the method of Massart (1987) under the condition that $\gamma < 1/4$. (Note that the parameter β in Andrews and Pollard (1994) is equal to 2γ in our notation.) Hence, we have enlarged the upper bound from 1/4 to 1/3. However it remains as an open problem if a further improvement is possible, and if further we can reach the upper bound 1 for strong mixing processes.

(c) To demonstrate that our general results lead to well-known (optimal) rates of convergence in special cases, we briefly discuss the case $\gamma = 0$. With $h = c_n \left(\frac{\sigma^2}{n}\right)^{\frac{1}{d+4}}$, where $c_n \to c > 0$ as $n \to \infty$, the results below follow from Theorem 2.3 immediately.

(c1) Let $\sigma^2 = 1$, we have that

$$n^{\frac{2}{d+4}} \sup_{C \in \mathcal{C}} |\widehat{F}_n(C|x) - F(C|x)| = O_P(1).$$

(c2) Let $\{\mathcal{C}_n\}$ be a sequence of classes of sets with $\mathcal{C}_n \subset \mathcal{C}$ and $\sup_{C \in \mathcal{C}_n} F(C|x) \leq \sigma_n^2 \leq 1$. Let $\sigma^2 = \sigma_n^2 \to 0$ and for which the conditions of Theorem 2.3 hold. Then

$$\left(\frac{n}{\sigma_n^2}\right)^{\frac{2}{d+4}} \sup_{C \in \mathcal{C}_n} |\widehat{F}_n(C|x) - F(C|x)| = O_P(\sqrt{\log n}).$$

3 Bahadur-Kiefer-type approximations

In this section we study the behaviour of the volume process defined in (1.5), which can be regarded as a generalized quantile process. Note that $\widehat{\mu}_{\mathcal{C}}(\alpha|x) = \mathrm{Leb}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x))$, and

$$\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x) \in \operatorname{argmax}\{\mathrm{Leb}(C) : \widehat{F}_n(C|x) \geq \alpha\}.$$

We assume throughout this section, that empirical MV-sets with finite ν-measure exist for every $\alpha \in [0,1]$. This assumption is satisfied for all standard choices of the class $\mathcal{C}$. Replacing the Lebesgue measure by a general function $\lambda : \mathcal{C} \to \mathbf{R}$, the process defined in (1.5) becomes a conditional version of the generalized quantile function as defined in Einmahl and Mason (1992). It reduces to the conditional quantile if we let $\mathcal{C} = \{(-\infty, x], x \in \mathbf{R}\}$ and $\lambda((-\infty, x]) = x$. In fact we have that the MV-set $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x) = (-\infty, \widehat{F}_n^{-1}(\alpha|x)]$ on the one hand, and the "volume" $\widehat{\mu}_{\mathcal{C}}(\alpha|x) = \lambda((-\infty, \widehat{F}_n^{-1}(\alpha|x)]) = \widehat{F}_n^{-1}(\alpha|x)$ on the other hand. Hence, a conditional quantile may be regarded as an MV-sets itself, and as well as its "volume".

A classical (unconditional) empirical MV-sets is the so-called *shorth* which is the MV-interval at the level 1/2. The term 'shorth' was first introduced by Andrews et al. (1972) referring to the *mean* of the data lying inside the MV-interval at the level 1/2, which is different from current practice. Rousseeuw (1986) introduced the MV-ellipsoid in the context of robust estimation for multivariate location and scatter.

A very important type of MV-sets are the so-called level sets defined in terms of probability density functions. Suppose that $F(\cdot|x)$ has Lebesgue density $g(\cdot|x)$. Denote

$$\Gamma_{g(\cdot|x)}(\lambda) = \{x \in \mathbf{R}^{d'} : g(\cdot|x) \geq \lambda\}, \quad \lambda > 0, \tag{3.1}$$

the *level sets* of $g(\cdot|x)$. It is easy to see that if $\Gamma_{g(\cdot|x)}(\lambda) \in \mathcal{C}$, it is an MV-set at the level $\alpha_\lambda = F(\Gamma_{g(\cdot|x)}(\lambda)|x)$.

Theorem 3.1 below presents Bahadur-Kiefer type rates of approximation for the set-indexed conditional empirical process. Note that $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ depends on the bandwidth h through $\widehat{F}_n(\cdot|x)$, which is not reflected explicitly in the notation.

Theorem 3.1 (Generalized Bahadur-Kiefer approximation)
Suppose that the conditions of Theorem 2.3 hold. Assume that $\mu_{\mathcal{C}}(\cdot|x)$ is differentiable with Lipschitz-continuous derivative $\mu'_{\mathcal{C}}(\cdot|x)$, and the condition (R_γ) holds for $\mathcal{C}$. Let further $\alpha \in (0,1)$ be fixed and suppose that $\mathrm{M}_{\mathcal{C}}(\alpha|x)$ is unique up to Leb-nullsets, that $F(\mathrm{M}_{\mathcal{C}}(\beta|x)|x) = \beta$ for all β in a neighborhood of α, and that $\mu'_{\mathcal{C}}(\alpha|x) > 0$. If for h and σ^2 satisfying the conditions of Theorem 2.3 *we have that as $n \to \infty$,*

$$d_{F(\cdot|x)}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x), \mathrm{M}_{\mathcal{C}}(\alpha|x)) = O_P(\sigma^2),$$

then as $n \to \infty$,

$$|(\widehat{F}_n - F)(\mathrm{M}_{\mathcal{C}}(\alpha|x)) + \frac{1}{\mu'_{\mathcal{C}}(\alpha|x)}(\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x))| = O_P\left(\frac{\Lambda_\gamma(\sigma^2, n)}{(nh^d)^{1/2}}\right).$$

In order to evaluate explicit rates from this theorem, we need to know the rates of convergence σ^2 for the empirical MV-sets. To this end, we assume that the level sets of the conditional density are (essentially) unique MV-sets. More precisely, it is assumed that for $\alpha \in [0,1]$ there exists a level λ_α such that for any $\mathrm{M}_{\mathcal{C}}(\alpha)$ we have

$$d_{\mathrm{Leb}}(\Gamma_{g(\cdot|x)}(\lambda_\alpha), \mathrm{M}_{\mathcal{C}}(\alpha|x)) = 0. \tag{3.2}$$

This assumption is fulfilled for all α if $\Gamma_{g(\cdot|x)}(\lambda) \in \mathcal{C}$ *for all* $\lambda \geq 0$, and $g(\cdot|x)$ has no flat parts (*i.e.* $\mathrm{Leb}\{y : g(y|x) = \lambda\} = 0 \;\; \forall\, \lambda > 0$). In addition we assume that $g(\cdot|x)$ is regular at the level λ_α, in the sense that

$$\mathrm{Leb}\{y : |g(y|x) - \lambda_\alpha| \leq \epsilon\} = O(\epsilon). \tag{3.3}$$

Under (3.2) and (3.3) rates of convergence for MV-sets are derived in Polonik and Yao (1998). Using these rates we obtain the following corollary.

Corollary 3.2 *Let conditions* (A2) – (A5), (B1) *and* (B2) *hold, and suppose that the process* $\{(X_t, Y_t)\}$ *is exponentially strong mixing. Let* $\alpha \in [0,1]$ *such that (3.2) and (3.3) hold. Then for* $\eta > 0$ *and*

$$h = c\ \max\left(n^{-\frac{1}{d+(3+\gamma)}}, n^{-\frac{1}{d+2(3\gamma+1)}}\right),$$

we have that as $n \to \infty$

$$|(\widehat{F}_n - F)(\Gamma_{g(\cdot|x)}(\lambda_\alpha)|x) + \lambda_\alpha(\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x))| = \begin{cases} O_P(n^{-\frac{2}{d+(3+\gamma)}+\eta}) & \text{if } \gamma < 1/5, \\ O_P(n^{-\frac{2}{d+2(3\gamma+1)}}) & \text{if } \gamma \geq 1/5. \end{cases}$$

Finally, we state a theorem giving Bahadur-Kiefer approximations for the more standard conditional one-dimensional empirical process indexed by $y \in \mathbf{R}$. Let

$$q(\alpha) = q(\alpha|x) = F^{-1}(\alpha|x)$$

denote the conditional quantile, and let

$$q_n(\alpha|x) = \widehat{F}_n^{-1}(\alpha|x),$$

where F^{-1} and $\widehat{F}_n^{-1}$ denote the generalized inverses of $F(\cdot|x)$ and $\widehat{F}_n(\cdot|x)$, respectively. Since we now use the optimal bandwidth, the bias comes into play (see also Lemma 5.2 in the Appendix). We define

$$\Psi_2(C|x) = \frac{1}{f(x)}\langle \nabla F(C|x), \int vK(v)\langle v, (\nabla f)(x)\rangle dv\rangle + \frac{1}{2}\int v^T\nabla^2 F(C|x)vK(v)\,dv \tag{3.4}$$

where ∇ and ∇^2 denote gradient and Hessian operator respectively. To simplify notation we write $\Psi_2(y|x)$ instead of $\Psi_2((-\infty, y]|x)$.

Corollary 3.3 (Bahadur-Kiefer approximation for the usual conditional empirical process) *Let conditions* (A2) – (A5), (B1) *and* (B2) *hold with* $\mathcal{C} = \{(-\infty, y], y \in \mathbf{R}\}$, *and suppose that the process* $\{(X_t, Y_t)\}$ *is exponentially strong mixing. Suppose further, that for a fixed* $\alpha \in (0,1)$ *the function* $g(\cdot|x)$ *is continuous at* $q(\alpha|x)$ *and that* $g(q(\alpha|x)|x) > 0$. *Let* $h = cn^{-1/d+4}$. *Then as* $n \to \infty$, *it holds almost surely that*

$$|(\widehat{F}_n - F)(q(\alpha|x)|x) + \Psi_2(q(\alpha|x)|x) + g(q(\alpha|x)|x)(q_n(\alpha|x) - q(\alpha|x))| = O\left(n^{-\frac{3}{d+4}}\sqrt{\log n}\right).$$

Remark 3.4 (a) Although the class $\mathcal{C}$ in Corollary 3.3 satisfies (R_γ) with $\gamma = 0$, the rates in Corollary 3.3 are faster than that derived from Corollary 3.2 with $\gamma = 0$. In fact, the quantiles converge at the rate of $1/\sqrt{nh^d}$, whereas the estimators of level sets converge slowerly, although both of them are MV-sets. Note that quantiles are MV-sets in the class of intervals of the form $(-\infty, y], y \in \mathbf{R}$, which have one fixed end-point at $-\infty$. Hence, the estimation of a quantile reduces to the estimation of its "length", which can be fulfilled at the rate of $1/\sqrt{nh^d}$. However, estimation of a general MV-set is much more involved, and hence the convergence is slower. (It is well-known that the classical shorth can be estimated at the rate of $n^{-1/3}$ only, whereas the length of the shorth can be estimated at the rate of $n^{-1/2}$.)

(b) Corollary 3.3 improves a result from Bhattacharya and Gangopadhyay (1990) which dealt with an *i.i.d.* case using a uniform kernel with one-dimensional X_i, i.e. $d = 1$. The convergence rate obtained by Bhattacharya and Gangopadhyay (1990) is $O\left(n^{-\frac{3}{5}}\log n\right)$, which is slower than ours by a factor $\sqrt{\log n}$.

(c) The above approximation rate is of the form $(nh^d)^{-3/4}\sqrt{\log n}$. Hence, up to a log-factor it is in alignment with the rates for unconditional (global) quantile process. For example, the almost sure rate for one-dimensional process (*i.e.* $d = 1$) with *i.i.d.* observations is $O(n^{-3/4}(\log\log n)^{3/4})$ (Kiefer 1967).

(d) The factor $g(q(\alpha)|x)$ in Theorem 3.3 corresponds to λ_α in Theorem 3.1. Note that both of them have the same geometric interpretation as the values of the (conditional) density at the boundary of the corresponding MV-set which are $\Gamma_{g(\cdot|x)}(\lambda_\alpha)$ and $(-\infty, q(\alpha|x)]$, respectively.

4 Discussion

Apart from its direct application in prediction of nonlinear and non-Gaussian time series (Polonik and Yao 1998, and references therein), a conditional empirical MV-set is also interesting (i) as an estimator for a level set of a conditional density, and (ii) to be used in tests for unimodality of conditional distributions. In this section, we discuss how the above theoretical results can be applied to these two applications.

First, we briefly illustrate how to derive the L_1-rate of convergence for a conditional empirical MV-set by applying Theorem 3.1 iteratively. It can be shown that the L_1-distance between the empirical and the true MV-set can be estimated from above by a sum of several terms including the difference of the empirical process and the generalized quantile process. (See Polonik and Yao 1998 for details.) Hence, Bahadur-Kiefer rates derived in Theorem 3.1 are useful. Note, however, an explicit rate σ^2 is needed in applying Theorem 3.1, and further, it is not necessary in Theorem 3.1 to let σ^2 converge to 0. Now we start with $\sigma^2 = 1$. Then Theorem 3.1 yields the first Bahadur-Kiefer approximation rate which in turn can be used to derive the first rate σ^2 for $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$. Further, this rate for $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ can be plugged into Theorem 3.1 to yield a faster Bahadur-Kiefer type approximation. This faster Bahadur-Kiefer rate leads to a faster rate of convergence for $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ and so on. The iteration will be continued until the rate of convergence is saturated.

The testing for modality of conditional distribution is an interesting and challenge problem in statistics. It has been observed that the conditional distribution of (nonlinear) time series given its lagged values could be multimodal. Further, the number of modes may vary over different places in the state space. Polonik and Yao (1998) proposed a heuristic device to detect the possible multimodality based on coverage probabilities of not necessarily connected regions. A more rigorous statistical test can be constructed as follows based on conditional MV-sets. We only consider a special case when Y is univariate (*i.e.* $d' = 1$).

To predict Y from X, the best predictive region among a candidate class $\mathcal{C}$ is the MV-set of $\mathcal{C}$ in the sense that the MV-set has the minimum Lebesgue measure. Obviously this best predictor depends on the choice of the class $\mathcal{C}$. In view of simple prediction, there is strong temptation to let $\mathcal{C}$ be the class of all intervals $\mathcal{I}_1$. However, such a $\mathcal{C}$ is only pertinent when the conditional density $g(\cdot|x)$ is unimodal. Indeed, if $g(\cdot|x)$ is, for example, bimodal, we should let $\mathcal{C} = \mathcal{I}_2$ which is the class of unions of at most two intervals. In this case, the MV-set of $\mathcal{I}_1$ may have much larger Lebesgue measure than that of $\mathcal{I}_2$. Hence, the comparison of the volumes (Lebesgue measures) for the MV-sets in different set classes gives us the information on the modality of the underlying conditional distribution. This idea has been explored by Polonik (1997) in testing the modality for unconditional distribution.

To test the null hypothesis that $g(\cdot|x)$ is unimodal, we define the statistic

$$T_{n,A}(x) = \sup_{\alpha \in A} (\hat{\mu}_{\mathcal{I}_2}(\alpha|x) - \hat{\mu}_{\mathcal{I}_1}(\alpha|x)) \tag{4.1}$$

where $A \subset [0,1]$. Obviously, we may replace $\mathcal{I}_1$ and $\mathcal{I}_2$ in the above expression by appropriate $\mathcal{C}$ and $\mathcal{D}$ (with $\mathcal{C} \subset \mathcal{D}$) respectively for testing different hypotheses. Now, it follows from Theorem 3.1 and its proof that under the null hypothesis

$$\hat{\mu}_{\mathcal{I}_2}(\alpha|x) = \mu_{\mathcal{I}_2}(\alpha|x) + \mu'_{\mathcal{I}_2}(\alpha|x)\Big((\hat{F}_n - F)(\Gamma_{g(\cdot|x)}(\lambda_\alpha)|x) + (nh^d)^{-1/2}\omega_{\nu_{n,\mathcal{I}_2}}(\sigma_n^2)\Big) + R_n,$$

where $\omega_{\nu_{n,\mathcal{I}_2}}$ denotes the modulus of continuity of $\nu_{n,\mathcal{I}_2}$ which is the conditional empirical process indexed by $\mathcal{I}_2$, and σ_n^2 denotes the L_1-rate of convergence of $\widehat{\mathrm{M}}_{\mathcal{I}_2}(\alpha|x)$ to $\Gamma_{g(\cdot|x)}(\lambda_\alpha)$. The remainder term R_n is of smaller order. The analogous expansion also holds for $\hat{\mu}_{\mathcal{I}_1}(\alpha|x)$. Since under the null hypothesis $\mu_{\mathcal{I}_2}(\alpha|x) = \mu_{\mathcal{I}_1}(\alpha|x)$, the statistic $T_{n,\{\alpha\}}(x)$ converges to 0 under the null hypothesis and the rate of convergence is $(nh^d)^{-1/2}\omega_{\nu_{n,\mathcal{I}_2}}(\sigma_n^2)$. The rates given in Corollary 3.2 for $\gamma = 0$ are explicit rates for this quantity for some particular h. Since the statistic $T_{n,A}$ is defined as a supremum, we need to show that the results in Theorem 3.1 and Corollary 3.2 hold uniformly for $\alpha \in A$, which can be validated at least for $A \subset [\epsilon, 1-\epsilon](\epsilon > 0)$ under appropriate conditions on the smoothness of $g(\cdot|x)$ (see Polonik 1997 for the global *i.i.d.* case).

The idea of excess mass provides an alternative approach to test the unimodality. The excess mass approach was introduced independently by Müller and Sawitzki (1987) and Hartigan (1987). (For further work see Nolan 1991, and Polonik 1995a,b). Adapted to the conditional empirical processes, the basic statistic is of the form $E_{n,\mathcal{C}}(\lambda|x) = \sup_{C \in \mathcal{C}}(\widehat{F}_n(C|x) - \lambda\,\mathrm{Leb}(C))$, which might be called a conditional empirical excess mass functional. As a function of λ it contains information about mass concentration of the underlying distribution. Similar to the above, we compare the functionals under different classes $\mathcal{C}$. Namely, we define the test statistic

$$T_n(x) = \sup_{\lambda > 0}(E_{n,\mathcal{I}_2}(\lambda|x) - E_{n,\mathcal{I}_1}(\lambda|x)),$$

which is a conditional version of the test statistic proposed by Müller and Sawitzki (1987, 1991).

The rates of convergence of $T_n(x)$ under the hypothesis of unimodality can be derived from Theorem 2.3. It can be shown that $\sqrt{nh}\,T_n(x)$ can be estimated from above by $\sup_{\lambda>0}(\nu_n(\Gamma_{n,\mathcal{I}_2}(\lambda)|x) - \nu_n(\Gamma_{g(\cdot|x)}(\lambda)|x))$. Here $\Gamma_{n,\mathcal{I}_2}(\lambda)|x) \in \mathcal{I}_2$ denotes the conditional empirical λ-cluster which is the maximizer of the excess mass statistic $E_{n,\mathcal{I}_2}(\lambda|x)$ defined above. See Polonik (1995a) for unconditional cases with *i.i.d.* observations. If L_1-rates of convergence σ_n^2 for the sets $\Gamma_{n,\mathcal{I}_2}(\lambda)$ to $\Gamma_{g(\cdot|x)}(\lambda)$

can be derived, then we have $\sqrt{nh}\, T_n(x) \le O_P(\omega_{\nu_{n,\mathcal{I}_2}}(\sigma_n^2))$, and rates of convergence of the quantity on the right-hand side of the last inequality immediately follow from Theorem 2.3.The rates σ_n^2 can be derived by using ideas from Polonik (1995) and the results of the present paper.

Finally, we point out that the above test can be generalized to tests for other null hypotheses if we replace $\mathcal{I}_1$ and $\mathcal{I}_2$ by $\mathcal{C}$ and $\mathcal{D}$ (with $\mathcal{C} \subset \mathcal{D}$) in the definition of the test statistic. This generalization can be treated analogously, provided information about the metric entropy with bracketing of $\mathcal{D}$ (and hence also about $\mathcal{C}$) is known. This shows the actual strength of Theorem 2.3.

5 Appendix: Proofs

Throughout the proofs we use the notation:

$$\varphi_n(C|x) = \frac{1}{nh^d} \sum_{t=1}^{n} I\{Y_t \in C\} K\left(\frac{X_t - x}{h}\right), \tag{5.1}$$

and define

$$f_n(x) = \frac{1}{nh^d} \sum_{t=1}^{n} K(\frac{X_t - x}{h}). \tag{5.2}$$

The corresponding theoretical functions are $\varphi(C|x) = F(C|x)f(x)$ and $f(x)$ itself. We write $K_h(y) = \frac{1}{h^d} K\left(\frac{y}{h}\right)$. Moreover, unless stated otherwise x is assumed to fixed such that $f(x) > 0$.

Let us first introduce two technical lemmas without proofs:

Lemma 5.1 *Suppose that f is continuous at x. Suppose further that f is bounded, and that* (2.4) *holds. Then we have $\forall\, C \in \mathcal{C}$ that as $n \to \infty$*

$$|E(\varphi_n(C|x)) - \varphi(C|x)| = o_P(1), \quad \textit{and} \quad |E(\widehat{F}_n(C|x)) - F(C|x)| = o_P(1).$$

If (2.4) *holds uniformly over $C \in \mathcal{C}$ so are the assertions.*

In the following lemma we give the exact asymptotic behaviour of the bias terms. Its proof consists of tedious, but straightforward calculations using Taylor expansions. Details are omitted.

Lemma 5.2 *Suppose that (A2) – (A5) hold. Let $\Psi_1(C|x) = \langle \nabla F(C|x), \int vK(v)\langle v, \nabla f(x)\rangle dv\rangle + \frac{1}{2} f(x) \int v^T \nabla^2 F(C|x) v K(v)\, dv \;+\; \frac{1}{2} F(C|x) \int v^T \nabla^2 f(x) v K(v)\, dv$ and let Ψ_2 as defined in* (3.4). *Then we have for each x as $n \to \infty$ that uniformly in $C \in \mathcal{C}$*

$$(i) \quad h^{-2}(E\varphi_n(C|x) - \varphi(C|x)) \to \Psi_1(C|x)$$

$$(ii) \quad h^{-2}(E\widehat{F}_n(C|x) - F(C|x)) \to \Psi_2(C|x).$$

Proof of Theorem 2.1: We use the following decomposition:

$$\widehat{F}_n(C|x) - F(C|x) = \frac{1}{f(x)}(\varphi_n(C|x) - \varphi(C|x)) - \frac{\widehat{F}_n(C|x)}{f(x)}(f_n(x) - f(x)). \tag{5.3}$$

¿From this it is easy to see that we only need to show that as $n \to \infty$

$$\sup_{C\in\mathcal{C}} |\varphi_n(C|x) - \varphi(C|x)| = o_P(1) \quad \text{and} \tag{5.4}$$

$$|f_n(x) - f(x)| = o_P(1). \tag{5.5}$$

(5.5) is well known to hold under the present conditions (cf. Bosq 1996). That for every fixed $C \in \mathcal{C}$ we have $|\varphi_n(C|x) - \varphi(C|x)| = o_P(1)$ can be shown by similar arguments, and is omitted here. We just show, how to conclude uniform consistency from this by using finite metric entropy with inclusion. Fix $\epsilon > 0$. For $C \in \mathcal{C}$ let C^*, C_* be a bracket for C, i.e. $C_* \subset C \subset C^*$ and $F(C^* \backslash C_*|x) < \epsilon$. There exist finitely many such sets. Since for $A \subset B$ we have $\varphi_n(A|x) \le \varphi_n(B|x)$ and also $\varphi(A|x) \le \varphi(B|x)$ it follows

$$\begin{aligned}
\sup_{C\in\mathcal{C}}(\varphi_n(C|x) - \varphi(C|x)) &\le \sup_{C\in\mathcal{C}}(\varphi_n(C^*|x)) - \varphi(C_*|x)) \\
&\le \sup_{C\in\mathcal{C}}(\varphi_n(C^*|x) - \varphi(C^*|x)) + \sup_{C\in\mathcal{C}}(\varphi(C^*|x) - \varphi(C_*|x)) \\
&= \sup_{C^*}(\varphi_n(C^*|x) - \varphi(C^*|x)) + \sup_{C\in\mathcal{C}} F(C^* \setminus C_*|x) f(x) \\
&\le \sup_{C^*}(\varphi_n(C^*|x) - \varphi(C^*|x)) + f(x)\, \epsilon.
\end{aligned} \tag{5.6}$$

An analogous lower bound holds with C^* replaced by C_*. Since the first term in the last line is a supremum over finitely many sets (for fixed $\epsilon > 0$) it follows from pointwise consistency of φ_n that this term is $o_P(1)$, and hence we finally obtain (5.4).

q.e.d.

Proof of Theorem 2.2: The proof of this theorem nowadays has become more or less standard. We just outline the main steps and for details we refer to Bosq (1996). Using the fact that under the given conditions $f_n(x)$ is consistent we have

$$\nu_n(C|x) = \sqrt{nh^d}\ \left(\frac{1}{f(x)}(\varphi_n(C|x) - \varphi(C|x)) - \frac{F(C|x)}{f(x)}(f_n(x) - f(x))\right)(1 + o_P(1)). \tag{5.7}$$

Using the notation $\tilde{b}_t(C) = I\{Y_t \in \mathcal{C}\}K(\frac{X_t - x}{h})$ we obtain

$$\nu_n(C|x) \quad = \quad \frac{1}{\sqrt{nh^d}} \sum_{t=1}^{n} W_{tn}(C|x) \tag{5.8}$$

$$+ \ \sqrt{nh^d}\ \frac{1}{f(x)}\Big(E\varphi_n(C) - \varphi(C)\Big) - \sqrt{nh^d}\ \frac{F(C|x)}{f(x)}\left(EK_h(X_t - x) - f(x)\right), \tag{5.9}$$

where $W_{tn}(C|x) = \frac{1}{f(x)}(\tilde{b}_t(C) - E\tilde{b}_t(C)) - \frac{F(C|x)}{f(x)}(K(X_t - x) - EK(X_t - x))$. It is well known that under the assumptions of the theorem the bias of f_n converges to zero at a rate h^2. The same holds for the bias of $\varphi_n(C)$ (Lemma 5.2). Hence, the assumptions on h assure that the terms in (5.9) are asymptotically negligible. It remains to show that the right-hand side in (5.8) is asymptotically normal with the given variance. To see this, the proof of Theorem 2.3 of Bosq (1996) can easily by adapted. [1] As for adapting the estimates given there one can use the fact that $\tilde{b}_t(C) \le K(\frac{X_t - x}{h})$. Calculation of the asymptotic variance-covariance matrix is lengthy but straightforward.

q.e.d.

Proof of Theorem 2.3: For the proof we adapt the chaining idea (well-known from empirical process theory) to the present situation, and use exponential inequalities for strong mixing processes that we take from Bosq (1996).

We start with the decomposition of $\nu_n(C|x)$ given in (5.7) above. Since under the present assumptions $\sqrt{nh^d}(f_n(x) - f(x)) = O_P(1)$ (*e.g.* Bosq 1996), the second summand of the main term in (5.7) is of the order $O_P(\sigma^2)$. It remains to show, that $\sqrt{nh^d}\sup_{C\in\mathcal{C}_\sigma}(\varphi_n(C|x) - \varphi(C|x))$ is of the desired order. To see this first note that $E\varphi_n(C|x) - \varphi_n(C|x)$ is of the (uniform) order $O_P(h^2)$. Hence, the assumption $\sqrt{nh^d}\, h^2 \le \Lambda_\gamma(\sigma^2, n)$ ensures that the bias-terms is of the required order. Therefore, with $\tilde{\nu}_n(C|x) = \sqrt{nh^d}\,(\varphi_n(C|x) - E\varphi_n(C|x)$ it remains to show that the assertion of the theorem holds with ν_n replaced by $\tilde{\nu}_n$.

The exponential inequality stated in the following lemma is used frequently in the sequel. It is heavily based on an exponential inequality which can be found in Bosq (1996).

Lemma 5.3 *Under the present assumptions for each $\epsilon > 0$ and each integer $r \in [1, n/2]$ there exist positive constants c, c_1, c_2 such that for $C \in \mathcal{C}_\sigma$ and large enough n*

$$\begin{aligned} P(|\tilde{\nu}_n(C|x)| > \epsilon) \le \quad & 4\exp\left(-\frac{\epsilon^2}{c\left(\sigma^2 + \sqrt{\frac{n}{h^d}}\,\frac{\epsilon}{r}\right)}\right) \\ & + \exp\left[-c_1\frac{n}{r} + c_2\left(\log r + (0 \vee \log\frac{n}{h^d\epsilon^2})\right)\right]. \end{aligned} \tag{5.10}$$

The proof of this lemma is given below. The remainder of the proof of Theorem 2.3 follows the lines of the proof of Theorem 2.3 of Alexander (1984) (and the corresponding Correction (1987)). Therefore some details are omitted.

Remember that for $\eta > 0$ and each $C \in \mathcal{C}$ there exist brackets $C_*, C^* \in \mathcal{C}$ with $C_* \subset C \subset C^*$

[1] In the proof of Theorem 2.3 of Bosq (1996) $r^{3/4}$ has to be replaced by $r^{5/4}$ in the formula preceding (2.40).

and $F(C_* \setminus C^*|x) < \eta$. Let $\mathcal{B}(\eta)$ denote a collection of brackets with (finite) minimal number of sets, such that $|\mathcal{B}(\eta)| = N_I(\eta, \mathcal{C}_\sigma, F(\cdot|x))$. By definition of H_γ we trivially have under $R(\gamma)$ that $\log|\mathcal{B}(\eta)| \leq H_\gamma(\eta)$.

Now, let $\delta_0 \geq \delta_1 \geq \ldots \geq \delta_N$ and $\eta_0, \eta_1, \ldots, \eta_N$ be positive real numbers defined below. For δ_j let $C_{j,*}, C_j^*$ denote the brackets for $C \in \mathcal{C}$ at the level δ_j^2, which means $C_{j,*} \subset C \subset C_j^*$ and $F(C_j^* \setminus C_{j,*}|x) \leq \delta_j^2$. Let further $\epsilon, B > 0$ such that

$$\sum_{j=0}^{N} \eta_j \quad \leq \quad \frac{\epsilon B}{8}, \tag{5.11}$$

then it is easy to see that:

$$\begin{aligned}
& P(\sup_{C \in \mathcal{C}_\sigma} |\tilde{\nu}_n(C|x)| > B) \\
\leq \quad & |\mathcal{B}(\delta_0^2)| \sup_{C \in \mathcal{C}_\sigma} P(|\tilde{\nu}_n(C|x)| > (1 - \frac{\epsilon}{4})B) \\
+ \quad & \sum_{j=0}^{N-1} |\mathcal{B}(\delta_j^2)||\mathcal{B}(\delta_{j+1}^2)| P(\sup_{C \in \mathcal{C}_\sigma} |\tilde{\nu}_n(C_{j,*}|x) - \tilde{\nu}_n(C_{j+1,*}|x)| > \eta_j) \\
+ \quad & P(\sup_{C \in \mathcal{C}_\sigma} |\tilde{\nu}_n(C_{N,*}|x) - \tilde{\nu}_n(C|x)| > \frac{\epsilon}{8} B + \eta_N) \\
=: \quad & (I) + (II) + (III).
\end{aligned} \tag{5.12}$$

Expressions (I) - (III) are now estimated separately. As for (I) we choose δ_0 to satisfy

$$H_\gamma(\delta_0^2) = \frac{1}{2\,c} \left(\frac{(1 - \frac{\epsilon}{4})^2 B^2}{\sigma^2 + \sqrt{\frac{n}{h^d}}\ \frac{(1-\frac{\epsilon}{4})B}{r_0}} \right)$$

with $r_0 = \frac{1}{\sigma^2}\sqrt{\frac{n}{h^d}}\ (1 - \frac{\epsilon}{4})B$ such that $H_\gamma(\delta_0) = \frac{(1-\epsilon/4)^2 B^2}{2\,c\,\sigma^2}$. Using the exponential inequality (5.10) with $r = r_0$ leads to

$$(I) \quad \leq \quad 4 \quad \exp\left[- \frac{(1 - \frac{\epsilon}{4})^2 B^2}{4\,c\,\sigma^2} \right] \tag{5.13}$$

$$+ \quad \exp\left[\frac{(1 - \frac{\epsilon}{4})^2 B^2}{4\,c\,\sigma^2} - c_1 \frac{\sqrt{n h^d}\ \sigma^2}{(1 - \frac{\epsilon}{4})^2 B^2} + c_2 \left(\log n + (0 \vee \log \frac{n}{h^d (1 - \frac{\epsilon}{4})^2 B^2}) \right) \right]. \tag{5.14}$$

Since r_0 has to lie between 1 and $n/2$ we obtain the following two conditions

$$B \geq (1 - \epsilon/4)^{-1}\, \sigma^2 \sqrt{\frac{h^d}{n}} \quad \text{and} \tag{5.15}$$

$$B \leq \tfrac{1}{2} \sigma^2 \sqrt{n h^d} \tag{5.16}$$

Now, (5.13) becomes small if B/σ becomes large. To get (5.14) small we need that for some $M > 0$ large enough $\frac{B^2}{\sigma^2} - \frac{\sqrt{nh^d}\,\sigma^2}{B^2} \leq -M \log n$. This is equivalent to the condition

$$B^4 + M\ B^2 \sigma^2 \log n \leq \sqrt{n h^d}\ \sigma^4. \tag{5.17}$$

As for the estimation of (II) define $s = \sqrt{\frac{\epsilon B}{4\sqrt{nh^d}}}$ and with δ_0 from above choose N and $\delta_j,\ j \geq 1$, as $\delta_{j+1} = s \vee \sup\{x \leq \frac{\delta_j}{2} :\ H_\gamma(x^2) \geq 2\, H_\gamma(\delta_j^2)\}$, and $N = \min\{j : \delta_j = s\}$. We only consider the case

$$s < \delta_0 \tag{5.18}$$

such that $N \geq 1$. This is the more difficult case. The case $s \geq \delta_0$ follows more easily by arguments analogous to the one given in the Correction of Alexander (1984). We choose for $j = 0, \ldots, N$

$$\eta_j = \sqrt{20}\, c\, \delta_j \sqrt{H_\gamma(\delta_{j+1}^2)}\,.$$

With this choice it is easy to see that $\sum_{j=1}^N \eta_j \leq \sqrt{20}\, c\, 2^{3/2} \int_s^{\delta_0} \sqrt{H_\gamma(x^2)}\ dx$. Hence, in view of condition (5.11) we require

$$B \geq M \int_s^{\delta_0} \sqrt{H_\gamma(x^2)}\ dx \tag{5.19}$$

for $M \geq M_0 > 0$. (5.10) is now applied to each summand of (II) separately. To that end we choose quantities $r_j, j = 0, \ldots, N-1$, analogously to r_0. Observing that $F(C_{j,*}\Delta C_{j+1,*}|x) \leq 2\, \delta_{j+1}^2$ we choose

$$r_j = \frac{1}{2\, \delta_{j+1}^2} \sqrt{\frac{n}{h^d}}\, \eta_j. \tag{5.20}$$

To apply (5.10) with $r = r_j$ we need $1 \leq r_j \leq n/2$. That $r_j \geq 1$ for large enough n can be seen easily. Since r_j is increasing in j it remains to assure that $1 \leq r_N \leq n/2$. This leads to the conditions

$$B \geq \frac{H_\gamma(s^2)}{\sqrt{nh^d}} \tag{5.21}$$

$$B \leq \frac{1}{2} n^{\frac{3}{2}} h^{-\frac{d}{2}} H(s^2). \tag{5.22}$$

Now, plugging the above quantities into (5.10) we obtain

$$\begin{aligned}
(II) \quad &\leq \quad 4 \sum_{j=0}^{N-1} \exp\left[4\, H_\gamma(\delta_{j+1}^2) - \frac{\eta_j^2}{4\, c\, \delta_{j+1}^2}\right] \\
&\quad + \sum_{j=0}^{N-1} \exp\left[4\, H_\gamma(\delta_{j+1}^2) - c_1 \frac{\sqrt{nh^d}\ \delta_j}{\sqrt{H_\gamma(\delta_{j+1}^2)}} + c_2 \left(\log n + (0 \vee \log \frac{n}{h^d \eta_j^2})\right)\right] \\
&\leq \quad 4 \sum_{j=0}^{N-1} \exp\left[-\, H_\gamma(\delta_{j+1}^2)\right] && (5.23) \\
&\quad + \sum_{j=0}^{N-1} \exp\left[4\, H_\gamma(\delta_{j+1}^2) - c_1 \frac{\sqrt{nh^d}\ \delta_j}{\sqrt{H_\gamma(\delta_{j+1}^2)}} + c_2 \left(\log n + (0 \vee \log \frac{n}{h^d s^2 H_\gamma(s^2)})\right)\right]. && (5.24)
\end{aligned}$$

Using the fact that $H_\gamma(\delta_{j+1}^2) \geq 2\, H_\gamma(\delta_j^2)$ the term in (5.23) can be shown to be (at least) of the same order as (5.13). As for the term (5.24) note that the assumptions assure that $\log \frac{n}{h^d s^2 H_\gamma(s^2)} =$

$O(\log n)$. In order to get (5.24) small we need that for all $j = 0, 1, \ldots, N-1$

$$4\, H_\gamma(\delta_{j+1}^2) - c_1 \frac{\sqrt{nh^d}\ \delta_j}{\sqrt{H_\gamma(\delta_{j+1}^2)}} + c_3 \log n \le -A_j(n) \tag{5.25}$$

for some real valued functions A_j such that $\sum_{j=1}^N \exp(-A_j(n)) < \infty$. Since the left-hand side of (5.25) is increasing in j it suffices to choose $A_N(n)$ satisfying (5.25), which means that we need

$$4\sqrt{H(s^2)}\ \Big(H(s^2) + c_3 \log n + A_N(n)\Big) \le c_1 \left(nh^d\right)^{\frac{1}{4}} \sqrt{B}\ , \tag{5.26}$$

satisfying in addition

$$N\ \exp(-A_N(n)) < \infty. \tag{5.27}$$

It remains to consider (III). Using Lemma 5.2 and arguments as in (5.6) we obtain

$$\begin{aligned}
\tilde\nu_n(C|x) \quad &\le \tilde\nu_n(C^*|x) + \sqrt{nh^d}\ \left(E\varphi_n(C^*|x) - E\varphi_n(C_*|x)\right) && (5.28)\\
&\le \sqrt{nh^d}\ \left(\varphi_n(C^*|x) - E\varphi_n(C^*|x)\right) + O\Big(\sqrt{nh^d}\ h^2\Big) + \sqrt{nh^d}\ f(x)\eta_n. && (5.29)
\end{aligned}$$

Analogously, we have an estimate of $\tilde\nu_n(C|x)$ from below by replacing C^* by C_* in (5.28) and (5.29). Hence, we obtain

$$\begin{aligned}
(III) \quad &\le P(\sup_{C\in\mathfrak{C}_\sigma} |\tilde\nu_n(C_{N,*}|x) - \tilde\nu_n(C_N^*|x)| > \frac{\epsilon}{8}B + \eta_N - \sqrt{nh^d}\ f(x)\delta_N - c_3\sqrt{nh^d}\ h^2)\\
&\le P(\sup_{C\in\mathfrak{C}_\sigma} |\tilde\nu_n(C_{N,*}|x) - \tilde\nu_n(C_N^*|x)| > \frac{\epsilon}{8}B + \eta_N - c_4\sqrt{nh^d}\ f(x)\delta_N)\\
&\le P(\sup_{C\in\mathfrak{C}_\sigma} |\tilde\nu_n(C_{N,*}|x) - \tilde\nu_n(C_N^*|x)| > \eta_N) && (5.30)
\end{aligned}$$

For the second inequality we used the fact that $h^2 = O(\delta_N)$ or equivalently

$$B^2 \ge M\, nh^{d+8} \tag{5.31}$$

for some $M > 0$. Hence, (III) can be treated as (II) above.

Now we consider the different cases of γ and check the above conditions on B. Below we frequently use M to denote a positive constant which has to be chosen appropriately (usually large enough), and which usually is different at different places.

As for $\gamma = 0$ we have $\int_s^{\delta_0} \sqrt{H_\gamma(x^2)}\ dx = O\left(\sqrt{\delta_0^2 \log\frac{1}{\delta_0^2}}\right)$. In view of (5.13) we make the Ansatz $B^2 = \sigma^2 D(\sigma^2)$ with $D(\sigma^2) \to \infty$ as $\sigma^2 \to 0$. Using (5.19) leads to the choice $B = M\sqrt{\sigma^2 \log\frac{1}{\sigma^2}}$. Note that here N can be chosen as $N = O(\log\log n)$, such that $A_N(n) = \log n$ is a valid choice. With these choices, all the conditions given above are satisfied under the present assumptions: Condition (5.15) is satisfied automatically for large enough n, and (5.16) holds if $\frac{\sigma^2}{\log\frac{1}{\sigma^2}} \ge \frac{4M}{nh^d}$. Further, (5.17) holds for large enough n if $\frac{\sqrt{nh^d}}{\log\frac{1}{\sigma^2}\log n} \to \infty$ as $n \to \infty$. (5.22) follows automatically, and (5.21) holds if $\frac{nh^d\sigma^2 \log\frac{1}{\sigma^2}}{(\log n)^2} > M > 0$. Inequality (5.31) follows from the assumption that $\sqrt{nh^{d+4}} = O(\Lambda_0(\sigma^2, n))$. Finally (5.26) follows from

$$\frac{nh^d\sigma^2 \log\frac{1}{\sigma^2}}{(\log n)^6} \to \infty \quad \text{as } n \to \infty,$$

which is the strongest condition.

As for $\gamma > 0$ a crucial condition again is (5.26). This condition can be seen to holds if $B \geq M\left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}}$. For $0 < \gamma < 1$ we have $\int_s^{\delta_0} \sqrt{H_\gamma(x^2)}\, dx = O\left(\delta_0^{1-\gamma}\right)$. Using the definition of δ_0 and (5.19) we obtain $B \geq M\,(\sigma^2)^{\frac{1-\gamma}{2}}$. Similarly we obtain for $\gamma = 1$ that $B \geq M \log n$, and for $\gamma > 1$ that $B \geq \left(nh^d\right)^{\frac{\gamma-1}{2\gamma}}$. Hence, we choose

$$B = M\left(\max\left(\left(\sigma^2\right)^{\frac{1-\gamma}{2}}, \left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}}\right)\right). \tag{5.32}$$

Note further that with this choice of B we may assume $N = O(\log n)$, such that again $A_N(n) = \log n$ is a valid choice, satisfying (5.27). With these choices all the above conditions on B are satisfied under the present assumptions. To see this first assume that $(\sigma^2)^{\frac{1-\gamma}{2}} \geq \left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}}$, such that $B = M\,(\sigma^2)^{\frac{1-\gamma}{2}}$. In this case (5.15) hold automatically for large n, and (5.16) follows from $\sigma^2 \geq M\left(nh^d\right)^{-\frac{1}{\gamma+1}}$. Further, (5.17) holds if $\sqrt{nh^d}\,(\sigma^2)^{2\gamma} \geq M \log n$. (5.21) reads as $B \geq \left(nh^d\right)^{-\frac{1-\gamma}{2(\gamma+1)}}$, which means here $\sigma^2 \geq M\left(nh^d\right)^{-\frac{1}{\gamma+1}}$. (5.22) reads as $B \leq n^{\frac{3+\gamma}{2(\gamma+1)}} h^{-\frac{d(1-\gamma)}{2(1+\gamma)}}$ which of course is satisfied here since $\sigma^2 \leq 1$. Last but not least, the assumption $s \leq \delta_0$ means $B \leq \left(nh^d\right)^{\frac{\gamma}{2(\gamma+2)}}(\sigma^2)^{\frac{1}{\gamma+2}}$. Pluggin in our choice of B again leads to $\sigma^2 \geq M\left(nh^d\right)^{-\frac{1}{1+\gamma}}$. Summing up, all these conditions are satisfied automatically if $(\sigma^2)^{\frac{1-\gamma}{2}} \geq \left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}}$. It remains to consider the case

$$\left(\sigma^2\right)^{\frac{1-\gamma}{2}} < \left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}}, \tag{5.33}$$

such that $B = M\left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}}$. Here (5.15) holds automatically for large n, because $\sigma^2 \leq 1$. Conditions (5.16) and (5.17) lead to lower estimates for σ^2 which do not conflict with (5.33). (5.21) and (5.22) are also seen easily to be satisfied. Finally inequality (5.31) follows for the same reasons as given above in the case $\gamma = 0$.

q.e.d.

Proof of Lemma 5.3: We use an exponential inequality for strong mixing processes given in Bosq (1996), Theorem 1.3. This is applied to $\tilde{\nu}_n(C|x) = \sqrt{nh^d}\,\frac{1}{n}\sum_{t=1}^n (b_t(C) - Eb_t(C))$, where as above $b_t(C) = I\{Y_t \in \mathcal{C}\}K_h(X_t - x)$. Since $|b_t(C)| \leq M_1 h^{-d}$, for some constant $M_1 > 0$ this gives for $\epsilon > 0$

$$P(|\tilde{\nu}_n(C|x)| > \epsilon) \;= P\left(|\varphi_n(C|x) - E\varphi_n(C|x)| > \frac{\epsilon}{\sqrt{nh^d}}\right)$$

$$\leq 4\ \exp\Big(-\frac{\epsilon^2 r}{8nh^d v^2(r)}\Big) + 22\ \left(1+\frac{4\ M_1\ n^{1/2}}{h^{d/2}\epsilon}\right)^{1/2} r\ \alpha\left(\left[\frac{n}{2r}\right]\right) \quad (5.34)$$

where here $[x]$ denotes integer part of x, $\alpha(\cdot)$ denotes the strong mixing coefficient defined in (2.1), and r is an integer with $1 \leq r \leq n/2$. Further,

$$v^2(r) = \frac{8r^2}{n^2}\sigma^2(r) + \frac{c_1}{2n^{1/2}h^{3d/2}}\ \epsilon$$

$$\text{and} \qquad \sigma^2(r) = O\left(s\ \mathrm{Var}(b_1(C)) + s\sum_{k=1}^{s-1}\mathrm{Cov}\ (b_1(C), b_{k+1}(C))\right),$$

where $s = n/2r$. The crucial point now is to get a good estimate for Cov $(b_1(C), b_{k+1}(C))$ and with that for $\sigma^2(r)$ for an appropriately chosen r.

Now, let p be from assumption (B1), and let $q \in \mathbf{N}$ be such that $\frac{1}{p}+\frac{1}{q} = 1$. Let further $p', q' \in \mathbf{N}$ also satisfying $\frac{1}{p'}+\frac{1}{q'} = 1$. Then it follows:

$$\begin{aligned}
&\mathrm{Cov}\ (b_1(C), b_{k+1}(C)) \leq \int I\{Y_1 \in C\} K_h(X_1 - x) K_h(X_{k+1} - x) dP \\
\leq\ & \left(\int \left(1\{Y_1 \in C\} K_h^{1/p'}(X_1 - x)\right)^{p'} dP\right)^{1/p'} \cdot \left(\int \left(K_h^{1/q'}(X_1 - x) K_h(X_{k+1} - x)\right)^{q'} dP\right)^{1/q'} \\
\leq\ & \left(F(C|x)f(x) + O(h^2)\right)^{1/p'} \cdot \left(\int f_k^p\right)^{1/q'p} \cdot \left(\int K_h^q(u-x)du\right)^{1/q'q} \left(\int K_h^{q'q}(u-x)du\right)^{1/q'q} \\
=\ & \left(F(C|x)f(x) + O(h^2)\right)^{1/p'} O\left(h^{-d/q' + d/q'q}\right) \cdot O\left(h^{-d+d/q'q}\right) \\
=\ & \left(F(C|x)f(x) + O(h^2)\right)^{1/p'} O\left(h^{-d(1+\frac{1}{q'}(1-\frac{2}{q})}\right) \\
=\ & (F(C|x))^{1/p'} O\left(h^{-db}\right)
\end{aligned}$$

where $b = 1 + \frac{1}{q'}(1-\frac{2}{q})$. Since $q < 2$ we have $b < 1$. Using this estimate and arguing as in Bosq (1996), proof of Lemma 2.1, we obtain that

$$\sum_{k=1}^{p-1}\mathrm{Cov}\ (b_1(C), b_{k+1}(C)) = o(h^{-db}(F(C|x))^{1/p'}\log n). \quad (5.35)$$

Furthermore, it is well known that $h^d\mathrm{Var}(b_1(C)) = F(C|x)f(x)\int K^2 + O(h^2)$, such that by choosing p' sufficiently large we obtain

$$\sigma^2(r) = O(sh^{-d}F(C|x)) = O(sh^{-d}\sigma_n^2). \quad (5.36)$$

This leads to the estimate

$$v^2(r) = O\left(s^{-1}h^{-d}\sigma_n^2 + \frac{\epsilon}{\sqrt{nh^{3d}}}\right). \quad (5.37)$$

Plugging this estimate into (5.34) we obtain for some constant $c > 0$

$$P(|\tilde{\nu}_n(C|x)| > \epsilon) \leq 4\exp\left(-\frac{\epsilon^2}{c\left(\sigma_n^2 + \sqrt{\frac{n}{h^d}}\ \frac{\epsilon}{r}\right)}\right) + 22\left(1 + 4\epsilon^{-1}\sqrt{\frac{n}{h^d}}\right)^{1/2} r\ \gamma^s. \quad (5.38)$$

Using $r \leq n/2$ we obtain for the last term the following estimate which completes the proof:

$$
\begin{aligned}
22\left(1+4\epsilon^{-1}\sqrt{\tfrac{n}{h^d}}\right)^{1/2} r\,\gamma^s &\leq \exp\left[-\frac{n}{2\,r}\log\frac{1}{\gamma}+\log\frac{n}{2}+\log 22+\frac{1}{2}\log\left(1+4\epsilon^{-1}\sqrt{\frac{n}{h^d}}\right)\right] \\
&\leq \exp\left[-c_1\frac{n}{r}+c_2\left(\log q+(0\vee\log\frac{n}{h^d\epsilon^2})\right)\right].
\end{aligned}
$$

q.e.d.

Proof of Theorem 3.1: Let $q_n(\alpha) = \sqrt{nh^d}\,\frac{f(x)}{\mu_k'(\alpha|x)}\Big(\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x)\Big)$. It is easy to see (using analogous arguments as in Polonik (1997), proof of Lemma 7.1) that on the set $B_n = \{d_{F(\cdot|x)}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x), \mathrm{M}_{\mathcal{C}}(\alpha|x)) < \sigma^2\} \cup \{|\alpha_n^{\pm} - \alpha| < \sigma^2\} \cup \{\alpha_n^{\pm} \in (0,1)\}$ we have

$$
\sqrt{nh^d}\;\frac{1}{\mu_k'(\alpha|x)}\Big(\mu_{\mathcal{C}}(\alpha_n^-|x)-\mu_{\mathcal{C}}(\alpha|x)\Big) \leq q_n(\alpha) \leq \sqrt{nh^d}\;\frac{1}{\mu_k'(\alpha|x)}\Big(\mu_{\mathcal{C}}(\alpha_n^+|x)-\mu_{\mathcal{C}}(\alpha|x)\Big) \quad (5.39)
$$

where

$$
\alpha_n^{\pm} = \alpha \pm ((\widehat{F}_n - F)(\mathrm{M}_{\mathcal{C}}(\alpha|x)|x) + \omega_{\nu_n(\cdot|x)}(\sigma^2). \quad (5.40)
$$

and $\omega_{\nu_n(\cdot|x)}(\epsilon) = \sup_{\{C,D\in\mathcal{C}:\, d_{F(\cdot|x)}(C,D)<\epsilon\}} |\nu_n(C|x) - \nu_n(D|x)|$ denotes the modulus of continuity of the conditional empirical process $\nu_n(C|x) = \sqrt{nh^d}\;(\widehat{F}_n(C|x) - F(C|x))$. ¿From (5.39) and (5.40) together with the fact that $P(B_n) \to 0$ as $n \to \infty$ the assertion follows by applying a one-term Taylor expansion.

q.e.d.

Proof of Theorem 3.3: First note that all the results proven above hold analogously if we choose $\mathcal{C} = \{(-\infty, y] : y \in \mathbf{R}\}$ and replace Leb by the function ν defined through $\nu((-\infty,y]) = y$.

A first application of Theorem 2.3 with $\sigma^2 = 1$ shows that $\sup_{\{y\in\mathbf{R}\}}\sqrt{nh^d}\;|\widehat{F}_n(y|x) - F(y|x)| = O_P(1)$ as $n \to \infty$ (cf. Remark 2.4, (c1)). ¿From this it follows that as $n \to \infty$

$$
|\widehat{F}_n(q_n(\alpha|x)|x) - F(q_n(\alpha|x)|x)| = O_P\left((nh^d)^{-1/2}\right),
$$

and since $\widehat{F}_n(q_n(\alpha|x)|x) = F(q(\alpha|x)|x) + o_P(1/nh^d)$ we obtain

$$
|F(q_n(\alpha|x)|x) - F(q(\alpha|x)|x)| = O_P\left((nh^d)^{-1/2}\right).
$$

Observing $|F(q_n(\alpha|x)|x) - F(q(\alpha|x)|x)| = d_F((-\infty, q_n(\alpha|x)], (-\infty, q(\alpha|x)])$, and applying Theorem 3.1 a second time, but now with $\sigma^2 = (nh^d)^{-1/2}$, gives the asserted *stochastic* rate by choosing $h = n^{-\frac{1}{d+4}}$. Note that here we also have to take into account the bias which for this choice of h does not vanish (cf. Lemma 5.2). The resulting rate also holds almost surely. This follows by applying Borel-Cantelli-Lemma, and becomes clear from the estimates derived in the proof of Theorem 3.1.

q.e.d.

References

Alexander K.S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041-1067, Correction *Ann. Probab.* **15** 428-430.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rodgers, W.H., and Tukey, J.W. (1972). Robust estimation of location: survey and advances. Princeton Univ. Press, Princeton, N.J.

Andrews W.K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Inst. Stat. Review* **62** 119-132

Bhattacharya, P.K. and Gangopadhyay, A.K. (1990): Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400-1415.

Bosq, D. (1996). Nonparametric statistics for stochastic processes. Lecture Notes in Statistics No. 110, Springer, New York.

Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré* **31** 393-427

Dudley, R.M. (1974). Metric entropy of classes of sets with differentiable boundaries. *J. Approx. Theorie* **10** 227-236.

Dudley, R.M. (1984). A course in empirical processes. *Ecole d'Ete de Probabilites de Saint Flour XII-1982, Lecture Notes in Math.* **1097** 1-142, Springer, New York.

Einmahl, J.H.J. and Mason, D.M. (1992). Generalized quantile processes. *Ann. Statist.* **20**, 1062-1078.

Gangopadhyay A.K. and Sen, P.K. (1993). Contiguity in Bahadur type representation of a conditional quantile and application in conditional quantile process. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift,* J.K. Gosh, S.K. Mitra, K.R. Parathasarathy and B.L.S. Prakasa Rao *(eds.)* 219 - 232, Wiley Eastern Limited, Publishers

Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267-270.

Horvath, L. (1988). Asymptotics of conditional empirical processes. *J. Multivariate Analysis* **26** 184 - 206

Hyndman, R.J. (1995). Forecast regions for non-linear and non-normal time series models. *Int. J. Forecasting*, **14**, 431-441.

Hyndman, R.J. (1996). Computing and graphing highest density regions. *Ameri. Statist.* **50**, 361-365.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323-1342.

Lientz, B.P. (1970). Results on nonparametric modal intervals. *SIAM J. Appl. Math.* **19** 356-366

Massart, P. (1987). Invariance principles for empirical processes: the weakly dependent case. PhD-Thesis, Paris-Sud

Mehra, K.L., Sudhakara Rao, M., and Upadrasta, S.P. (1991). A smooth conditional quantile estimator and related applications of conditional empirical processes. *J. Multivariate Anal.* **37** 151-179

Müller, D.W. and Sawitzki, G. (1987). Using excess mass estimates to investigate the modality of a distribution. Preprint No. 398, SFB 123, Universität Heidelberg

Müller, D.W. and Sawitzki, G. (1991). Excess mass estimates and tests of multimodality. *J. Amer. Statist. Assoc.* **86** 738-746

Nolan, D. (1991). The excess mass ellipsoid. *J. Multivariate Anal.* **39** 348-371

Polonik, W. (1995a). Measuring mass concentration and estimating density contour clusters - an excess mass approach. *Ann. Statist.* **23** 855-881

Polonik, W. (1995). Density estimation under qualitative assumptions in higher dimensions. *J. Multivariate Anal.* **55** 61- 81

Polonik, W. (1997). Minimum volume sets and generalized quantile processes. *Stoch. Processes and Appl.* **69** 1-24

Polonik, W. and Yao, Q. (1998). Conditional minimum volume predictive regions for stochastic processes. (Submitted.)

Shorack G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

Stute, W. (1986a). Conditional empirical processes. *Ann. Statist.* **14** 638-647.

Stute, W. (1986b). On almost sure behaviour of conditional empirical distribution functions. *Ann. Probab.* **14** 891-901.

van der Vaart A. and Wellner J.A. (1996) *Weak convergence and empirical processes.* Springer, New York.

Xiang, X. (1995). On Bahadur-Kiefer representation of a kernel conditional quantile estimator. *Nonparametric Statistics* **5** 275-287

Xiang, X. (1996). A kernel estimator of a conditional quantile. *J. Multivariate Anal.* **59** 206-216